CDF Note 8677

Search for Electroweak Single-Top-Quark Production using Neural Networks with 955 pb^{-1} of CDF II data

The CDF Collaboration
URL http://www-cdf.fnal.gov
(Dated: January 29, 2007)

We report on a search for electroweak single-top-quark production with CDF II data corresponding to 955 pb^{-1} of integrated luminosity. We apply neural networks to construct discriminants that distinguish between single-top and background events. Two analyses are performed, assuming a top quark mass of 175 GeV/c^2. In the first one we combine t- and s-channel events to one single-top signal under the assumption that the ratio of the two processes is given by the standard model (SM). Using ensemble tests, we determine that we expect with a probability of 50% to see a single-top signal that is larger than a 2.6 σ fluctuation of the background (p-value of 0.5%). A binned likelihood fit to the data yields no evidence for single-top. The observed p-value is 54.6% and indicates that the data are compatible with the background hypothesis only. A combined single-top cross section above 2.6 pb is excluded at the 95% confidence level.
In the second analysis we separate the two single-top production modes. A binned likelihood fit to a two-dimensional distribution of two neural network outputs yields most probable values for the cross sections of $0.2^{+1.1}_{-0.2}$ pb for the t-channel and $0.7^{+1.5}_{-0.7}$ pb for the s-channel. The separate search analysis features an expected p-value of 0.4% (2.7 σ). The observed p-value, i.e. the probability for the data to be due to a background fluctuation only, is found to be 21.9%.

I. INTRODUCTION

According to the standard model, in $p\bar{p}$ collisions at the Tevatron top quarks can be created in pairs via the strong force, of singly via the electroweak interaction. The latter production mode is referred to as "single-top-quark" production and takes place mainly through the s- or t- channel exchange of a W boson. The CDF and DØ collaborations have published single-top results at $\sqrt{s} = 1.8$ TeV and $\sqrt{s} = 1.96$ TeV [1, 2]. None of these analyses established single-top evidence, and 95% confidence level (C.L.) upper limits on the single-top production cross section were set.
The theoretical single-top production cross section is $\sigma_{s+t} = 2.9 \pm 0.4$ pb for a top mass of 175 GeV/c^2 [3]. Despite this small rate, the main obstacle in finding single-top is in fact the large associated background. After all section requirements are imposed, the signal to background ratio is approximately 1/20. This challenging, background-dominated dataset is the main motivation for using multivariate techniques.

II. COMMON EVENT SELECTION

The CDF event selection exploits the kinematic features of the signal final state, which contains a top quark, a bottom quark, and possibly additional light quark jets. To reduce multijet backgrounds, the W originating from the top quark is required to have decayed leptonically. One therefore demands a single high-energy electron or muon ($E_T(e) > 20$ GeV, or $P_T(\mu) > 20$ GeV/c) and large missing transverse energy $\not\!\!E_T > 25$ GeV from the undetected neutrino.
The backgrounds belong to the following categories: $Wb\bar{b}$, $Wc\bar{c}$, Wc, mistags (light quarks misidentified as heavy flavor jets), top pair production $t\bar{t}$ events (one lepton or two jets are lost due to detector acceptance), non-W (QCD multijet events where a jet is erroneously identified as a lepton), $Z \to ll$ and diboson WW, WZ, and ZZ. We remove a large fraction of the backgrounds by demanding exactly two jets with $E_T > 15$ GeV and $|\eta| < 2.8$ be present in the event. At least one of these two jets has to be tagged as a b quark jet by using displaced vertex information from the silicon vertex detector (SVX) of CDF [4]. The non-W content of the selected electron dataset is further reduced by several requirements to the angle between the $\not\!\!E_T$ vector and the transverse momentum vector of the jets. The numbers of expected and observed events are listed in figure 1.

CDF II Preliminary 955 pb^{-1}

W + 2 jets	
$Wb\bar{b}$	170.9 ± 50.7
$Wc\bar{c}$	63.5 ± 19.9
Wc	68.6 ± 19.0
Mistags	136.1 ± 19.7
QCD multijet	26.2 ± 15.9
$t\bar{t}$	58.4 ± 13.5
Diboson	13.7 ± 1.9
Z + jets	11.9 ± 4.4
Total Background	549.3 ± 95.2
t-channel	22.4 ± 3.6
s-channel	15.4 ± 2.2
Total Prediction	587.1 ± 96.6
Observation	644

FIG. 1: Expected number of signal and background events and total number of events observed in 955 pb^{-1} in the CDF single-top dataset.

III. FIT TO B TAG NEURAL NETWORK

To cross-check the background estimate, we perform a fit to the output of a neural network b tagger. The network tagger is applied to jets that are already tagged by the secondary vertex tagger of CDF [4]. In case of double-tagged events the leading b jet (highest in E_T) is included in this distribution. The network output is quite characteristic, not only for b jets, but also for charm and light jets. The tagger thereby allows to determine the flavor composition of our data sample.

We create templates of the neural network output distributions for b, c and light jets using simulated events. Those templates are fitted to the W+jets data output distributions in the 1, 2 and 3 jets bin. The results of the fits are displayed in figure 2. For all three cases, the fitted distributions describe the data very well.

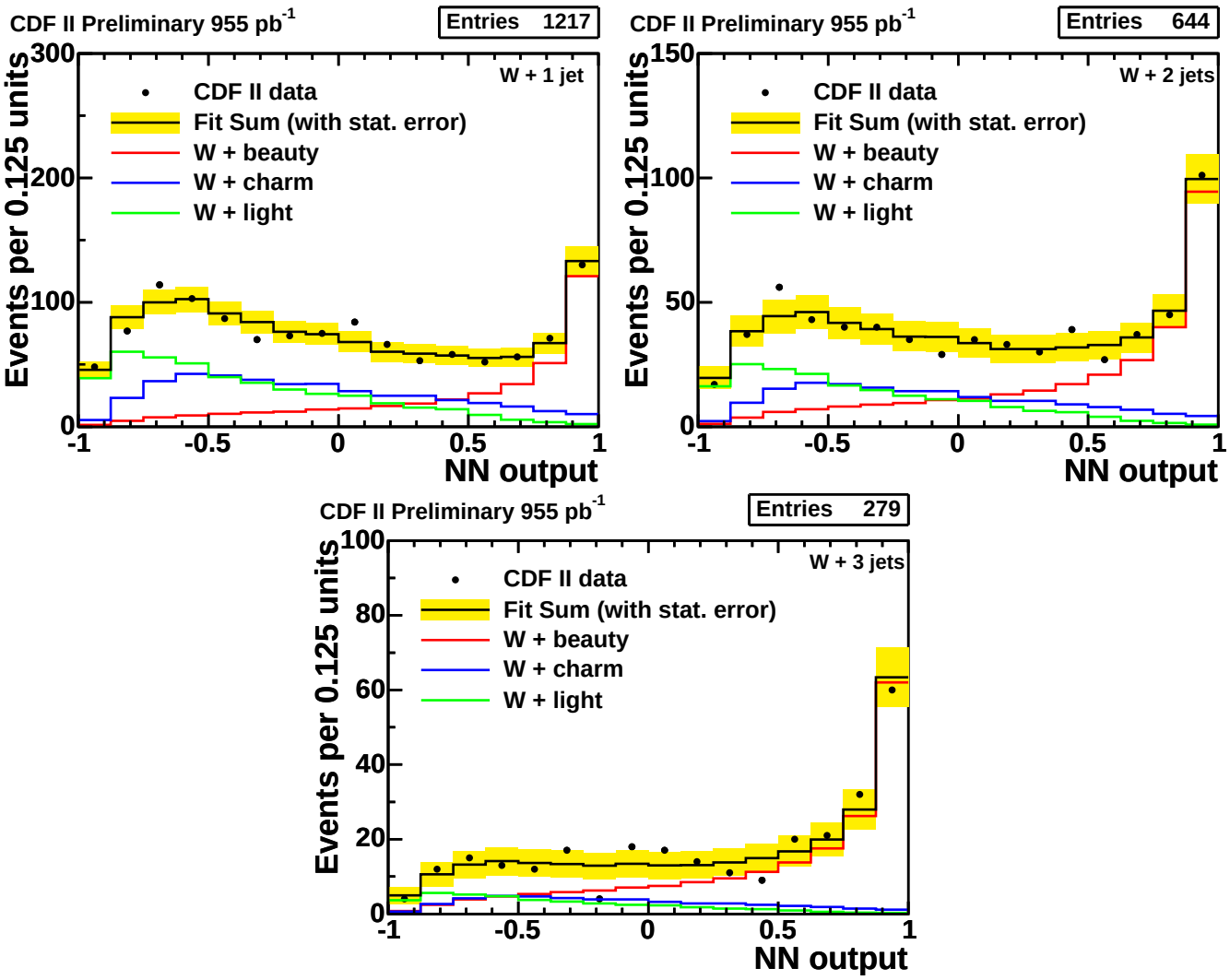

FIG. 2: Fit to the output of the neural network b tagger: The upper left plot shows the fit to the 1 jet bin, the upper right one the fit to the 2 jet bin and the lower one the fit to the 3 jet bin. The b templates are displayed in red, the charm ones in blue and the light ones in green. The sum of the fitted templates is shown in black with a yellow error band. The black points are the data distribution.

IV. NEURAL NETWORK INPUT VARIABLES

Using neural networks 26 kinematic or event shape variables are combined to a powerful discriminant. One of the variables is the output of a neural net b tagger (see figure 3). The neural net b tagger gives an additional handle to reduce the large background components where no real b quarks are contained, mistags and charm-backgrounds. Both of them amount to about 50% in the W+2 jets data sample even after imposing the requirement that one jet is identified by the secondary vertex tagger of CDF [4].

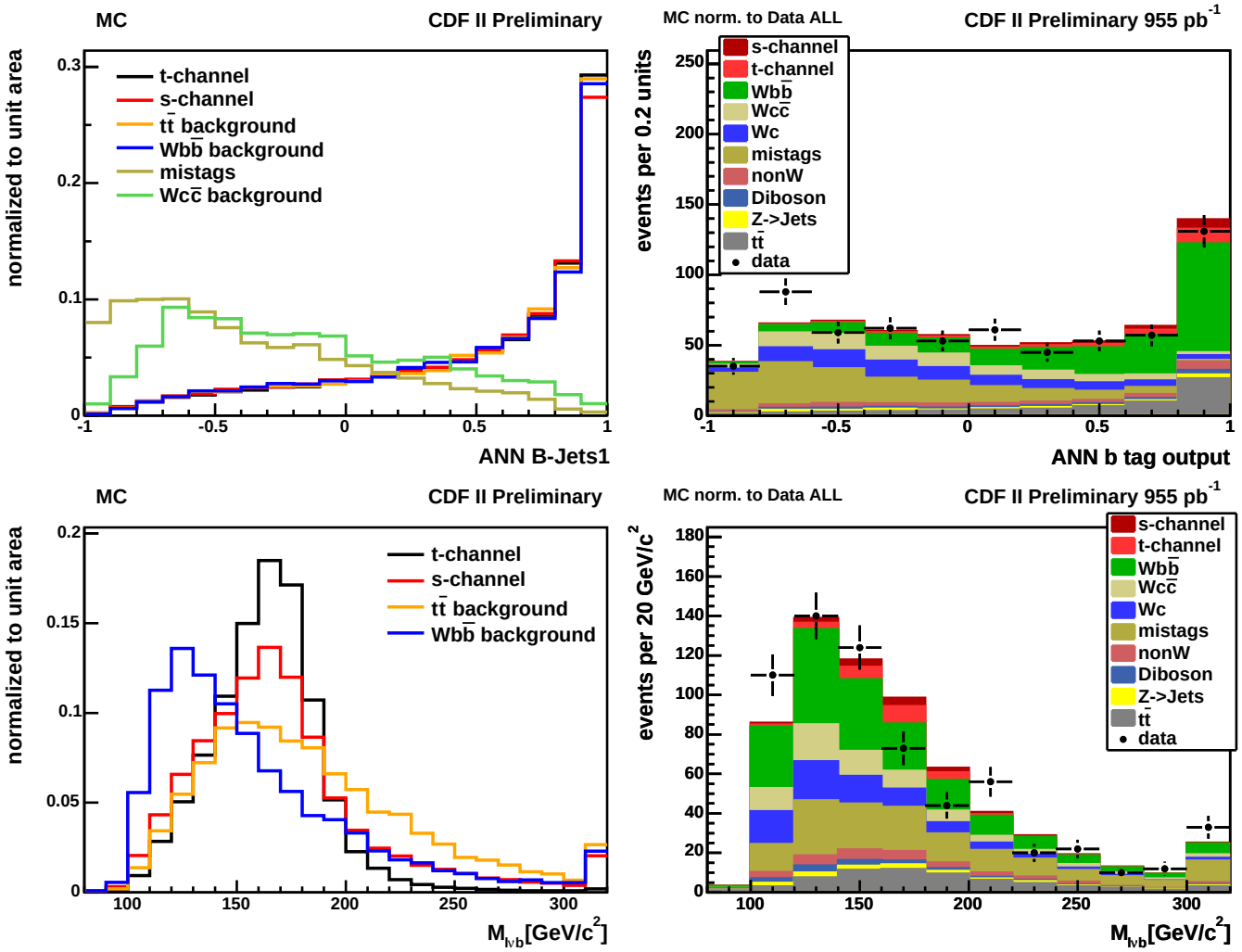

FIG. 3: Some of the most powerful neural networks input variables: MC distribution (left) and data - MC comparison (right) of the neural net b tagger (top row); MC distribution (left) and data - MC comparison (right) of the reconstructed top mass (bottom row).

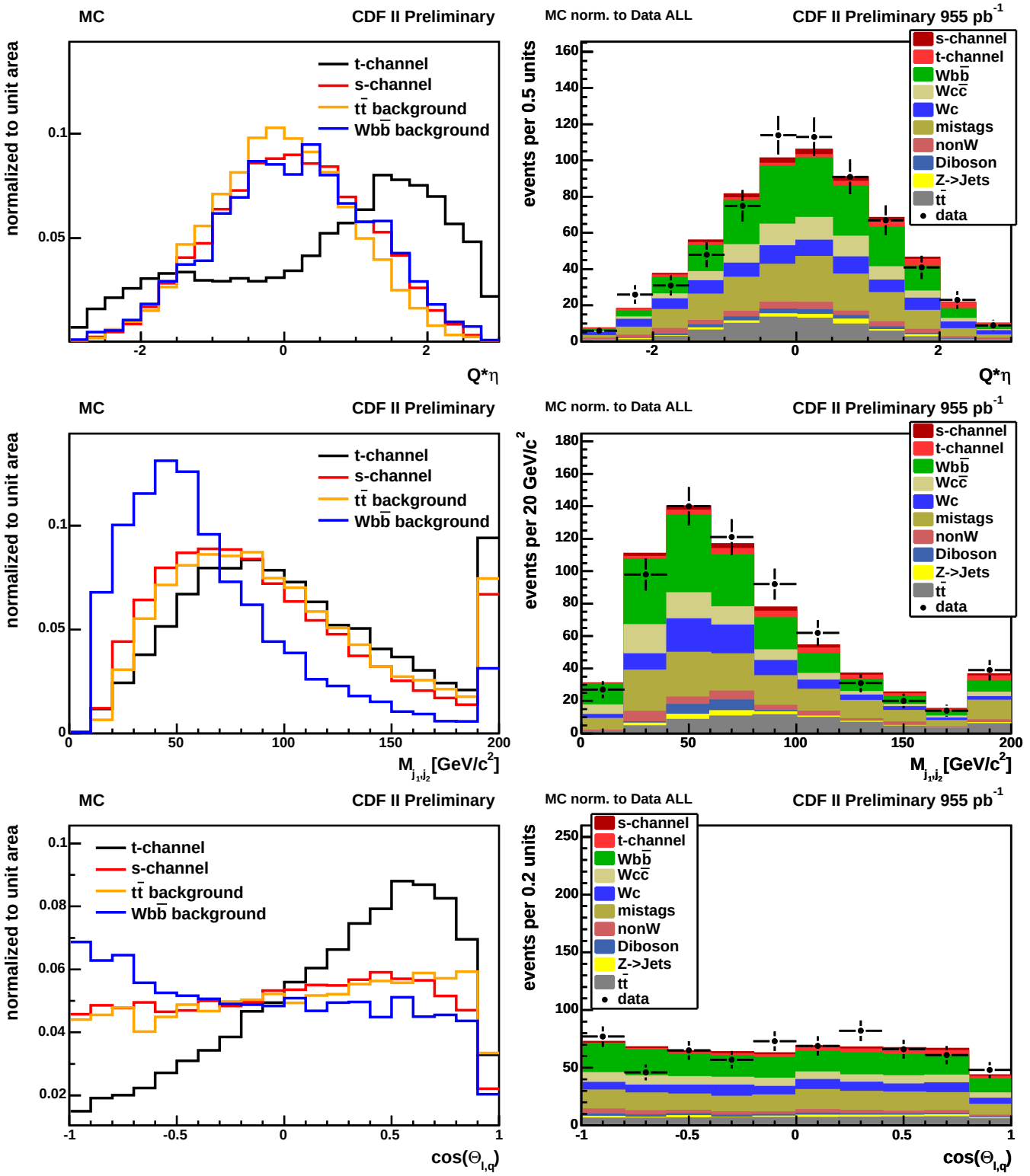

FIG. 4: Some of the most powerful neural networks input variables: MC distribution (left) and data - MC comparison (right) of the charge of the lepton times pseudorapidity of the leading light jet (top row); MC distribution (left) and data - MC comparison (right) of the invariant mass of the two leading jets (middle row); MC distribution (left) and data - MC comparison (right) of the cosine of the angle $\Theta_{l,q}$ where $\Theta_{l,q}$ is reconstructed by determining the angle between the tight lepton and the beam axis in the top rest frame (bottom row).

V. TEMPLATES FOR BINNED LIKELIHOOD FIT

A. Templates for Combined Search

In principle, it would be possible to create separate templates for each of the expected physical processes. Since it is difficult for a likelihood fit to distinguish between distributions that are not very distinct, it is more practical to combine similar shapes into one discriminant.
Comparing the neural network outputs for the simulated non-top backgrounds it can be seen that some of the distributions are very similar. Therefore, three non-top templates are created: $Wb\bar{b}$ and WZ build the so-called b-like background template. The c-like template consists of $Wc\bar{c}$, Wc, WW, and mistagged light events. The non-W background is combined with $Z \to ee$, $Z \to \mu\mu$, $Z \to \tau\tau$, and ZZ. Together with the $t\bar{t}$ template and the single-top processes, we get five templates for the combined search which are displayed in figure 5.

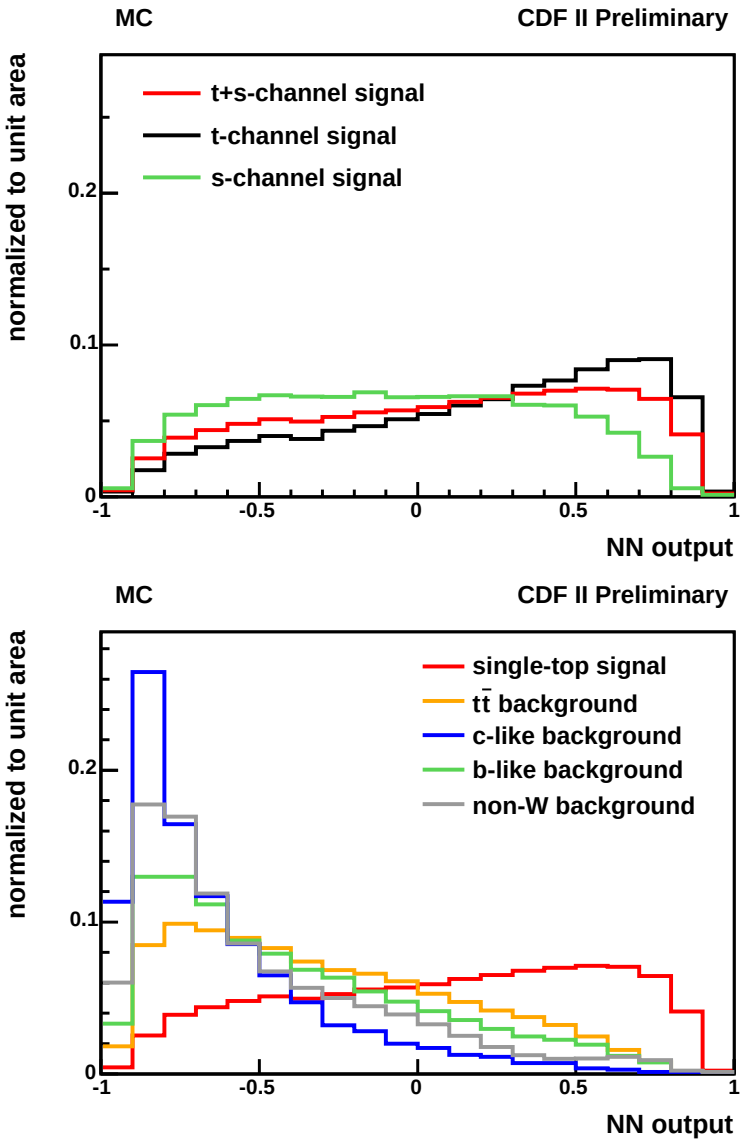

FIG. 5: s- and t-channel MC distributions are combined at SM ratio to the single-top template (top); the five MC templates for the combined search: single-top, $t\bar{t}$, c-like, b-like, and non-W (bottom).

B. Templates for Separate Search

For the separate search we use two independent neural networks, one trained for s-channel and the other one for t-channel, which provide the opportunity to search for both channels simultaneously. The creation of the templates for signal and background processes is made in a similar way as in the combined search, even though it is done in 2D for both network outputs simultaneously. With the two single-top signal templates (figure 6) and four background templates (figure 7), we use six discriminants in the separate search.

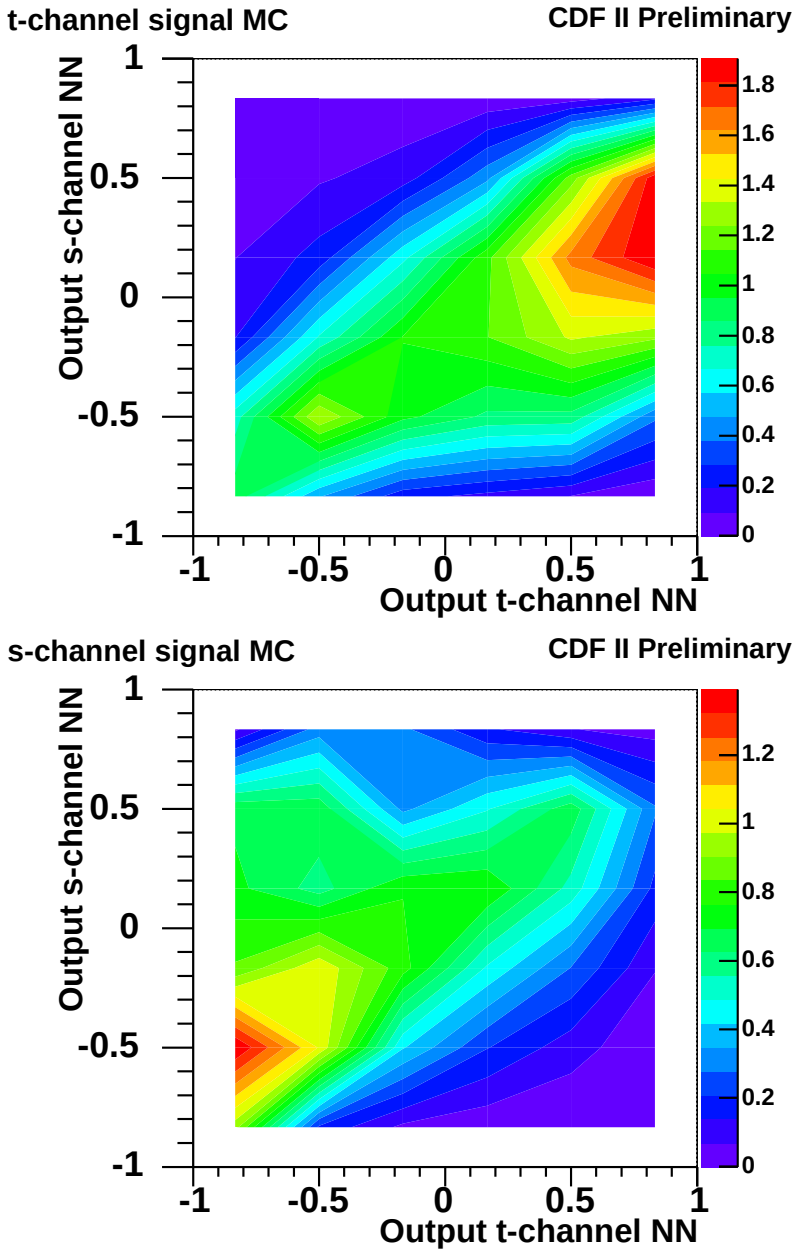

FIG. 6: 2D template of the t-channel signal MC (top); 2D template of the s-channel signal MC (bottom).

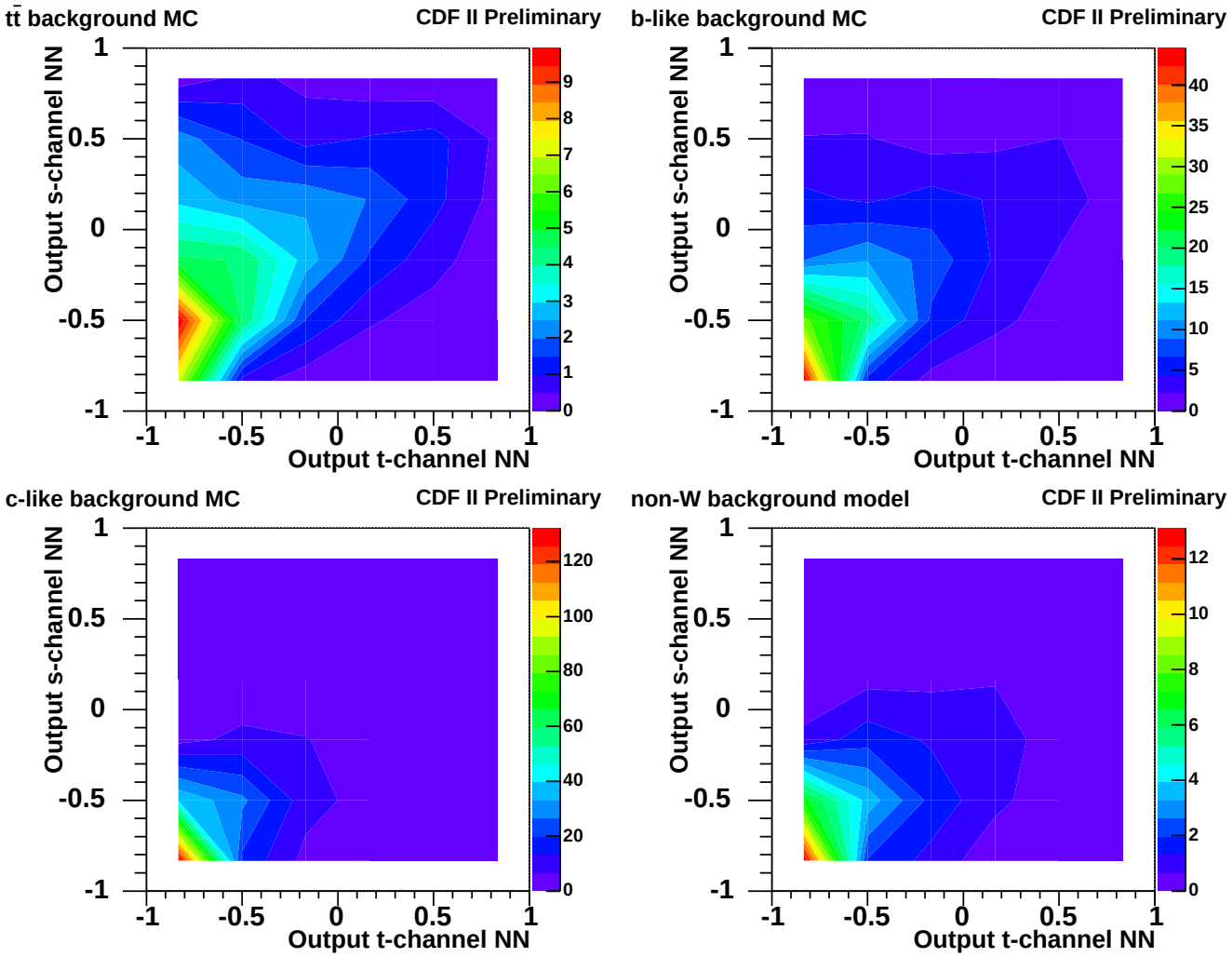

FIG. 7: 2D template of the $t\bar{t}$ background MC (top row left); 2D template of the b-like background MC (top row right); 2D template of the c-like background MC (bottom row left), 2D template of the non-W background model (bottom row right).

VI. SYSTEMATIC UNCERTAINTIES

Systematic uncertainties can cause a shift in the event detection efficiency for events of different physics processes, but can also cause a change in the shape of the template distributions.
The rate uncertainties are summarized in figure 8. Ten sources of systematic shape uncertainties are considered: the jet energy scale (JES), initial state gluon radiation (ISR), final state gluon radiation (FSR), parton distribution functions (PDFs), neural net b tagger, the factorization and renormalization scale for W + heavy flavor processes, the modeling of the W + heavy flavor samples, the modeling of mistag events, the flavor composition and modeling of non-W events.
The shape uncertainties are determined by altering the respective effects within their uncertainties. In this way two shifted distributions are obtained for first five sources (see three examples in figure 8), one plus and one minus distribution. For the last five systematic sources one alternative model is considered. Therefore, only one systematic shape is obtained for theses effects.

CDF II Preliminary 955 pb^{-1}

Systematic Rate Uncertainties				
Source	t-channel	s-channel	single-top	$t\bar{t}$
JES	± 2.1 %	± 1.4 %	± 1.8 %	± 8.6 %
ISR	± 2.6 %	± 1.6 %	± 2.2 %	± 10.1 %
FSR	± 3.4 %	± 1.4 %	± 2.6 %	± 11.4 %
PDF	± 2.5 %	± 2.2 %	± 2.4 %	± 2.4 %
MC	± 2.0 %	± 1.0 %	± 1.6 %	± 3.0 %
ϵ_{evt}	± 8.1 %	± 6.3 %	± 7.4 %	± 7.1 %
Luminosity	± 6.0 %	± 6.0 %	± 6.0 %	± 6.0 %
Cross section	± 13.0 %	± 13.0 %	± 13.0 %	± 19.7 %
M_{top} (not applied)	± 2.9 %	± 2.7 %	± 2.8 %	

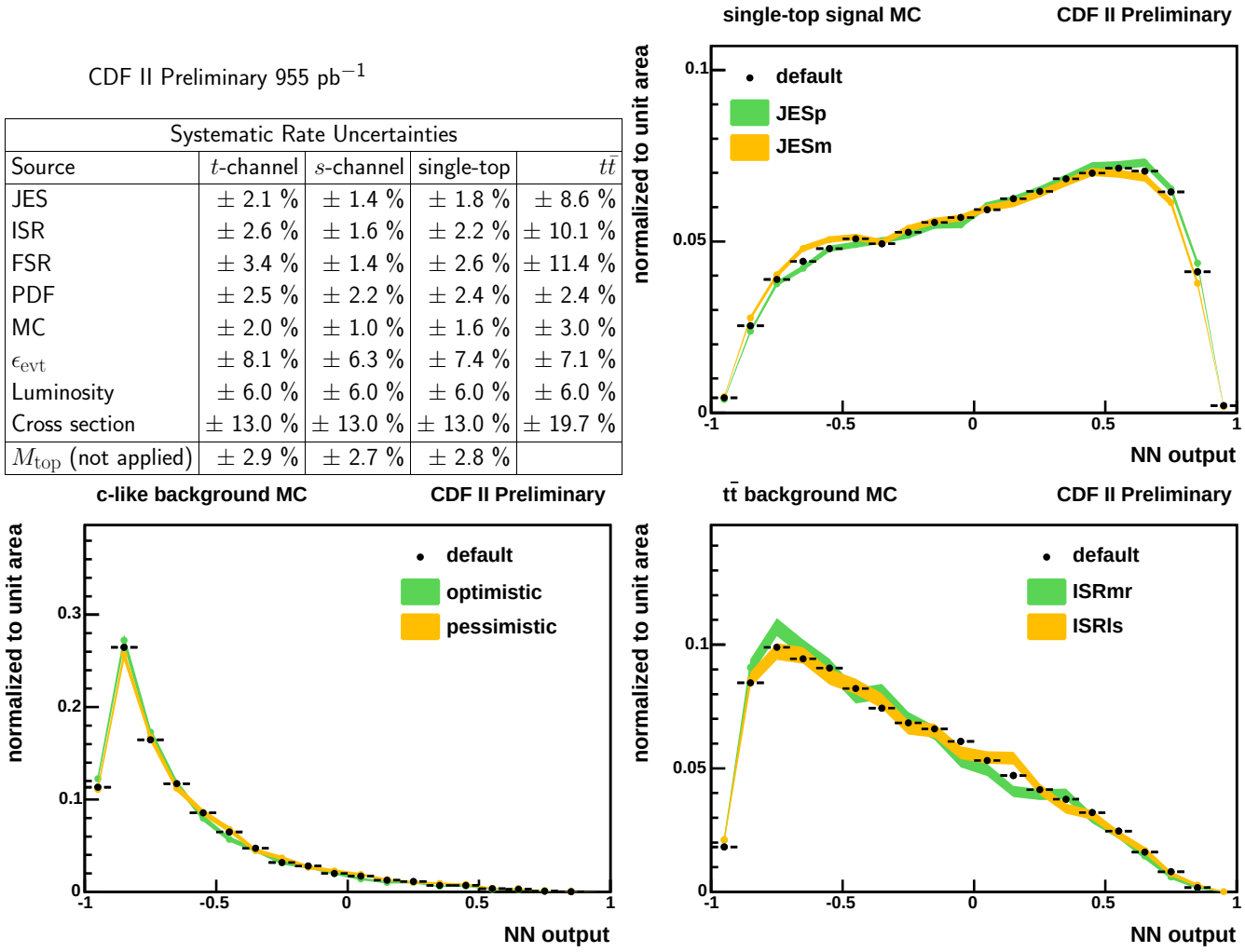

FIG. 8: Table of systematic rate uncertainties (top left); Single-top MC distribution of the systematic shape uncertainty due to the JES (top right); c-like MC distribution of the systematic shape uncertainty due to the neural net b tagger (bottom left); $t\bar{t}$ MC distribution of the systematic shape uncertainty due to the ISR (bottom right)

VII. LIKELIHOOD FUNCTION

The likelihood function consists of Poisson terms for the individual bins of the fitted histograms, Gaussian constraints to the background rates, and Gaussian constraints to the strengths of systematic effects.

$$L(\beta_1, ..., \beta_5; \delta_1, ..., \delta_S) = \prod_{k=1}^{B} \frac{e^{-\mu_k} \cdot \mu_k^{n_k}}{n_k!} \cdot \prod_{j=2}^{5} G(\beta_j, 1.0, \Delta_j) \cdot \prod_{i=1}^{S} G(\delta_i, 0.0, 1.0); \tag{1}$$

Systematic uncertainties are included as factors modifying the expectation value μ_k of events in a certain bin k.

$$\mu_k = \sum_{j=1}^{5} \beta_j \cdot \nu_j \cdot \mathcal{L}_{\text{int}} \cdot \left\{ \prod_{i=1}^{S} (1 + \delta_i \cdot \epsilon_{ji}) \right\} \cdot \alpha_{jk} \cdot \left\{ 1 + \sum_{i=1}^{S} (\delta_i \cdot \kappa_{jik}) \right\} \tag{2}$$

The index j runs over the different physics processes that occur in the likelihood function. The cross section of process j is σ_j. In the likelihood function we use the parameter β_j, which is the cross section normalized to its standard model prediction. The event detection efficiency of process j is named ν_j. The normalized content of bin k of the template histogram for process j is α_{jk}.
We consider five effects which cause systematic uncertainties in acceptance. Ten sources of uncertainties in the template shape are taken into account. The sources of systematic uncertainties are indexed with i. The relative acceptance uncertainties due to these sources are named ϵ_{ji}. The relative uncertainties in the bin content of bin k of the template histograms are called κ_{jik}. The variation in strength of a systematic effect i is measured with the variable δ_i.
The shape uncertainties are calculated from the systematically shifted histograms α_{jik}^{+} and α_{jik}^{-} according to

$$\kappa_{jik} = \frac{\alpha_{jik}^{+} - \alpha_{jik}^{-}}{2\,\alpha_{jk}} \tag{3}$$

That means, the systematic shifts are symmetrized. By construction the κ_{jik} satisfy the normalization condition

$$\sum_{k=1}^{B} \kappa_{jik} = 0\ . \tag{4}$$

The systematically shifted template that takes into account the shifts caused by all systematic effects with strengths $\{\delta_i\}$ is given by

$$\alpha'_{ji} = \alpha_{jk} \cdot \left\{ 1 + \sum_{i=1}^{S} \delta_i \cdot \kappa_{jik} \right\} \tag{5}$$

Due to (4) the shifted histogram α'_{ji} is properly normalized:

$$\sum_{k=1}^{B} \alpha'_{ji} = 1\,. \tag{6}$$

The background rates (cross sections) and the parameters describing the strength of systematic excursions (δ_i) are constrained by additional Gaussian terms in the likelihood. The background rates are constraint within the uncertainties of the prediction, Δ_j. The strengths of the systematic effects are constraint to 0.0 with a standard deviation of 1.0. The single-top content (cross section) is measured by fitting the parameters (β_j and δ_j) of the likelihood function to the observed data. This is achieved by minimizing the log likelihood with respect to these parameters using the program MINUIT.

Using this technique one can compute the likelihood as a function of the single-top cross section, β_1, only. The log likelihood is minimized at a fixed value of β_1 with respect to all other variables which are also often called *nuisance* parameters. The resulting one-dimensional function is called the reduced likelihood, $\mathcal{L}_{\text{red}}(\beta_1)$. This method is often called *profiling* the likelihood function.

VIII. EXPECTED SENSITIVITY AND SIGNIFICANCE

A. Combined Search

We use ensemble tests to compute the sensitivity of our analysis. An ensemble test consists of a set of pseudo experiments. For each pseudo experiment we determine first the number of events N_j of each process by drawing a random number from a Poisson distribution with a mean μ_j. In a second step we draw random numbers from the template distributions of the neural network output.
We perform two ensemble tests: one with single-top events included at the predicted standard model rate (see figure 9 at the left hand side), one without any single-top events (see figure 9 at the right hand side). The main results of the ensemble tests are shown below, where the most probable values for the rates of the different processes, i.e. the central values obtained from the likelihood fit for each pseudo experiment, are given in units of the expected rates. We define the RMS of single-top distribution as the expected uncertainty for a potential measurement of the cross section. We find a value of 45%. This figure includes all systematic uncertainties.

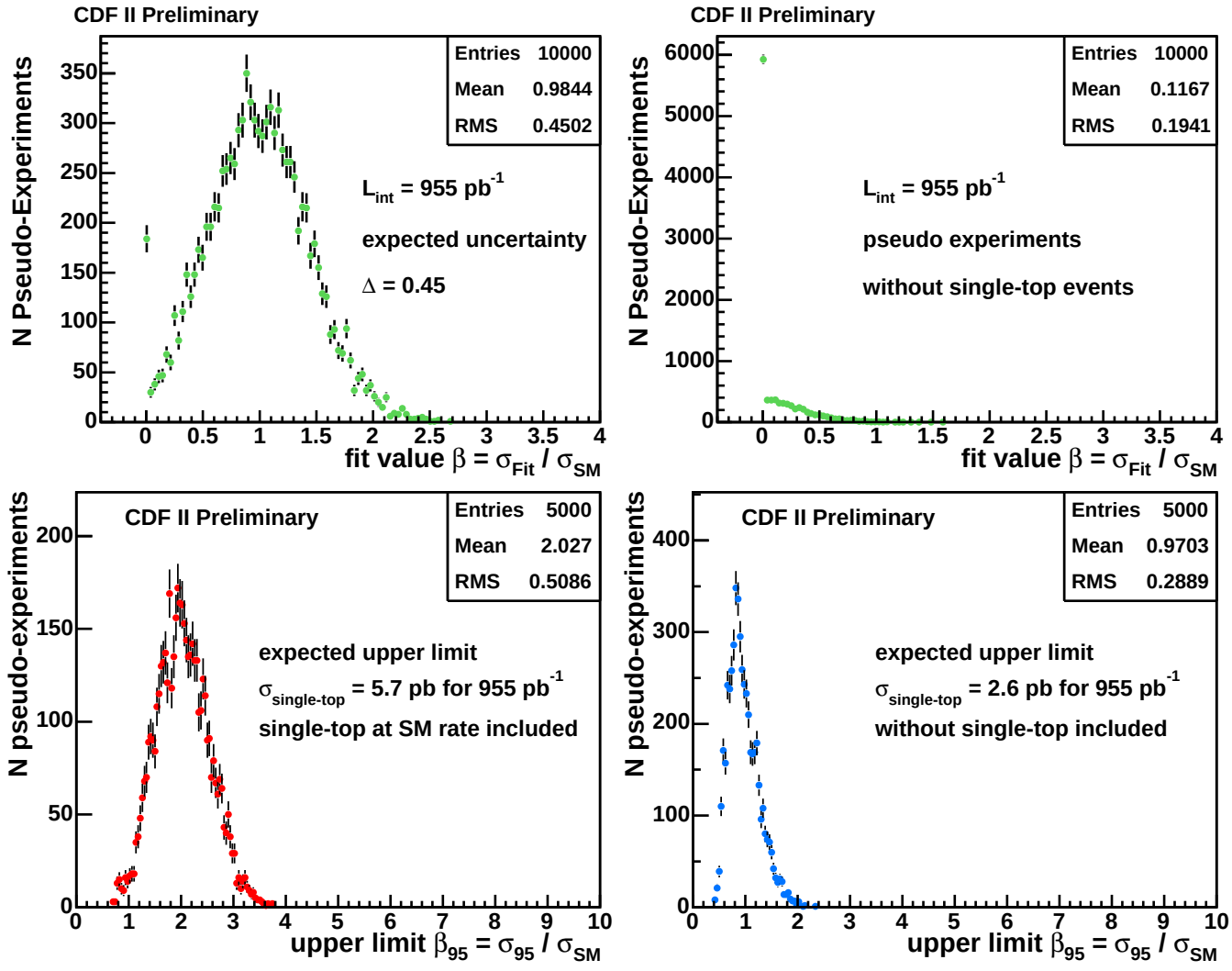

FIG. 9: Results of ensemble tests (top): distribution of expected single-top measurements. Left: single-top events are included in the pseudo experiments at the expected standard model rate; right: pseudo experiments are done without single-top events. Distribution of expected upper limits (bottom). Left: single-top events are included at the expected standard model rate. Right: the pseudo experiments include no single-top events. We define the median of the distribution as the expected upper limit.

To compute the significance of a potentially observed signal, we perform a hypothesis test. Two hypotheses are considered. The first one, H_0, assumes that the single-top cross section is zero ($\beta_1 = 0$) and is called the *null hypothesis*. The second hypothesis, H_1, assumes that the single-top production cross section is the one predicted by the standard model ($\beta_1 = 1$). The objective of our analysis is to observe single-top, that means to reject the null hypothesis H_0.

The hypothesis test is based on the Q-value,

$$Q = -2\left(\ln L_{\rm red}(\beta_1 = 1) - \ln L_{\rm red}(\beta_1 = 0)\right) , \tag{7}$$

where $L_{\rm red}(\beta_1 = 1)$ is the value of the reduced likelihood function at the standard model prediction and $L_{\rm red}(\beta_1 = 0)$ is the value of the reduced likelihood function for a single-top cross section of zero. Using the two ensemble tests the distribution of Q-values is determined for the case with single-top included at the standard model rate, q_1, and for the case of zero single-top cross section, q_0. The two Q-value distributions are shown in figure 10.

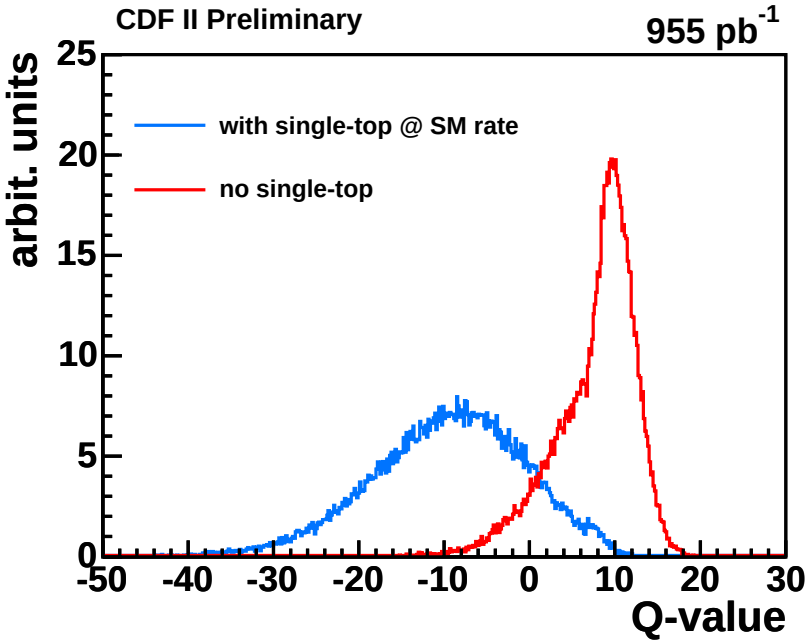

FIG. 10: Distributions of Q-values for two ensemble tests, one with single-top events present at the expected standard model rate, one without any single-top events.

In order to quantify the probability for the null hypothesis H_0 to be correct we define the p-value. When we assume that the value Q_0 is observed in a particular experiment, the p-value is given by

$$p(Q_0) = \int_{-\infty}^{Q_0} q_0(Q')\, dQ' \ , \tag{8}$$

where q_0 is the distribution of Q-values for the null hypothesis H_0. Often the p-value is also named $1 - {\rm CL}_b$.

To quantify the sensitivity of our analysis we define the expected p-value $\hat{p} = p(Q_1^{\rm med})$ where $Q_1^{\rm med}$ is the median of the Q-value distribution q_1 for the hypothesis H_1. The meaning of $\hat{p}$ is the following: Under the assumption that H_1 is correct one expects to observe $p < \hat{p}$ with a probability of 50%. We find $\hat{p} = 0.5\%$, including all systematic uncertainties. In other words, *a priori* we expect with a probability of 50% to reject the null hypothesis with 99.5% confidence level, under the assumption that the cross section has the predicted value. Yet a different wording: Assuming the predicted single-top cross section we expect, with a probability of 50%, to see at least that many single-top events that the observed excess over the background corresponds to a $2.6\,\sigma$ background fluctuation.

B. Separate Search

As for the combined search, we perform two sets of ensemble tests for the separate search: one with single-top events included at the predicted standard model rate (see figure 11 at the left hand side), one without any single-top events (see figure 11 at the right hand side). The main results of the ensemble tests are shown below, where the most probable values for the rates of the different processes, i.e. the central values obtained from the s- and t-channel likelihood fit for each pseudo experiment, are given in units of the expected rates.

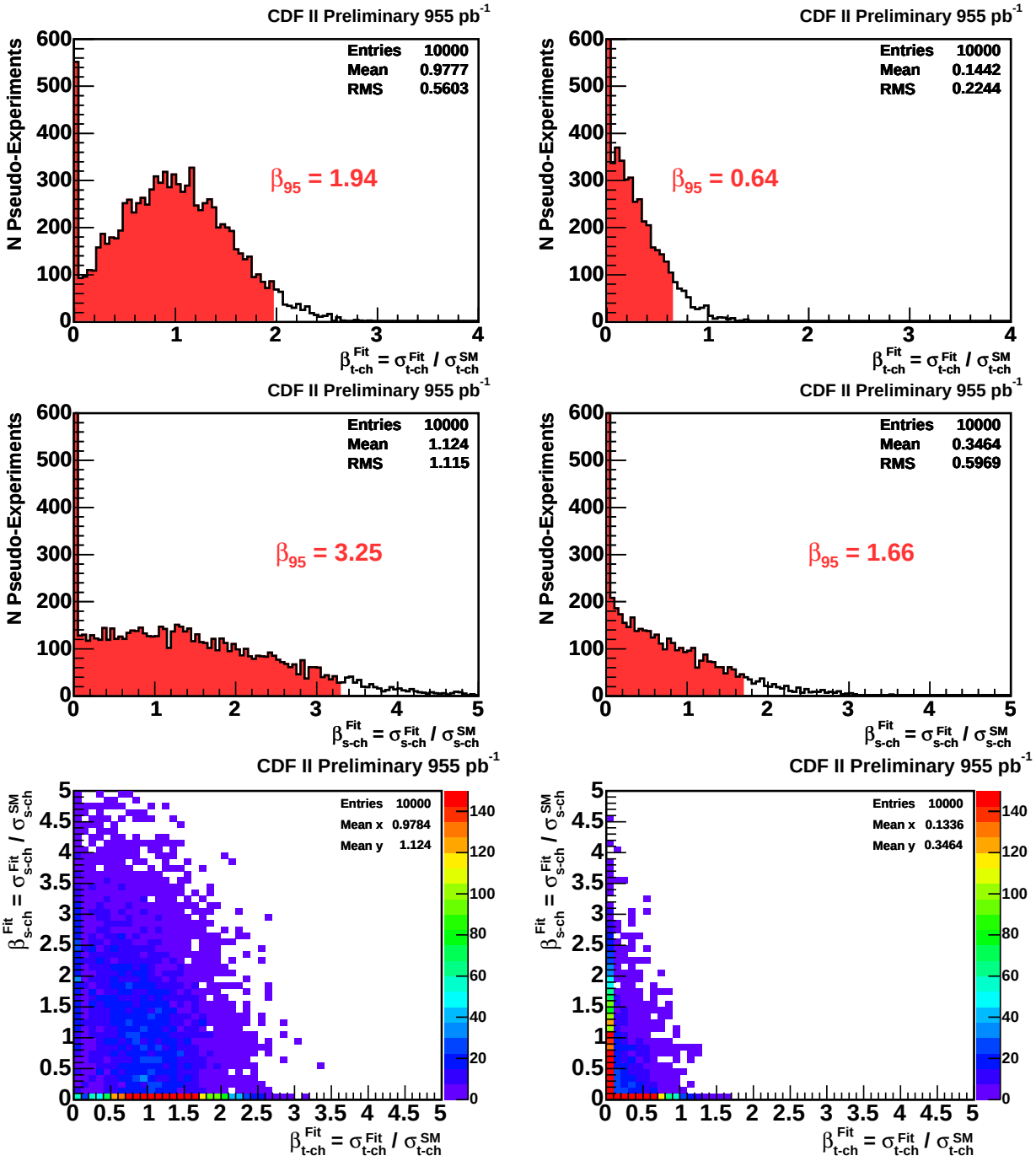

FIG. 11: Pseudo experiment distribution with (left hand side) and without (right hand side) single-top at SM rate: t-channel (top row) and s-channel (middle row) cross section measurement normalized to the SM prediction and the 2D distribution (bottom row). The expected t- and s-channel limits at 95% confidence level are $\sigma_{t-ch}^{95exp.} = 3.8$ pb and $\sigma_{s-ch}^{95exp.} = 2.9$ pb.

To obtain the significance of a potentially observed signal, we calculate the Q-value, as described above for the combined search. The two Q-value distributions (with and without SM single-top) are shown in figure 12. We find an expected p-value of 0.4% including all systematic uncertainties. Assuming the predicted single-top cross section, we expect, with a probability of 50%, to see at least that many single-top events that the observed excess over the background corresponds to a 2.7 σ background fluctuation.

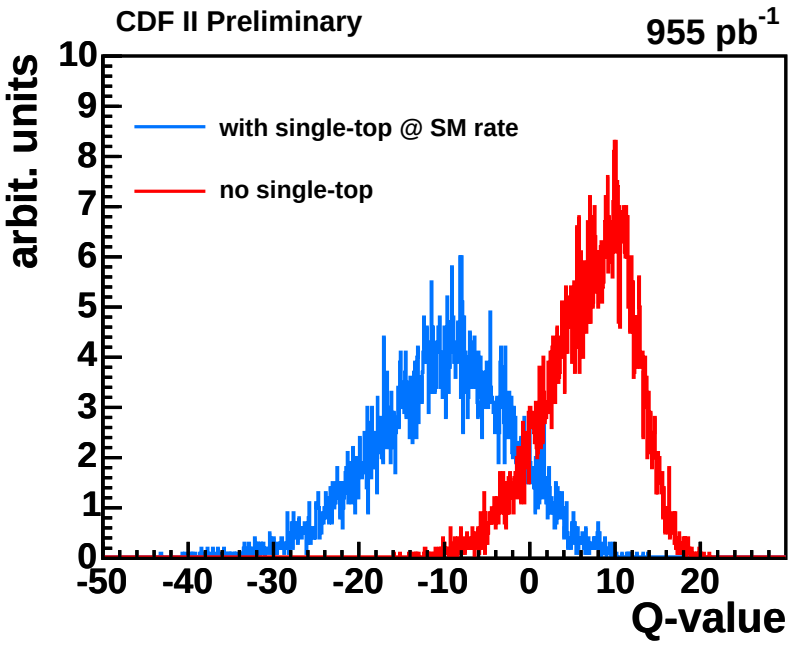

FIG. 12: Distributions of the Q-values for ensemble test with and without single-top present at SM rate.

IX. BINNED LIKELIHOOD FIT TO DATA FOR COMBINED SEARCH

In the signal region we expect 45.6 ± 7.5 events, while we observe 31 events in data. The data are displayed in figure 13.

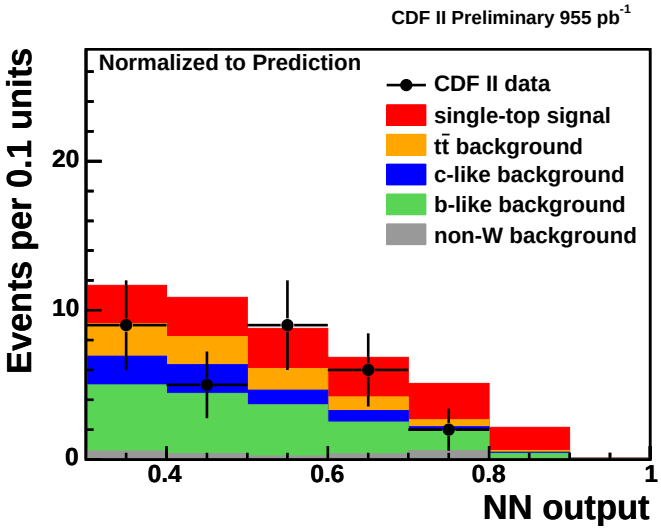

FIG. 13: Data distribution of the neural network output in the signal region.

The likelihood fit to the entire NN output region yields a rate of zero single-top events.

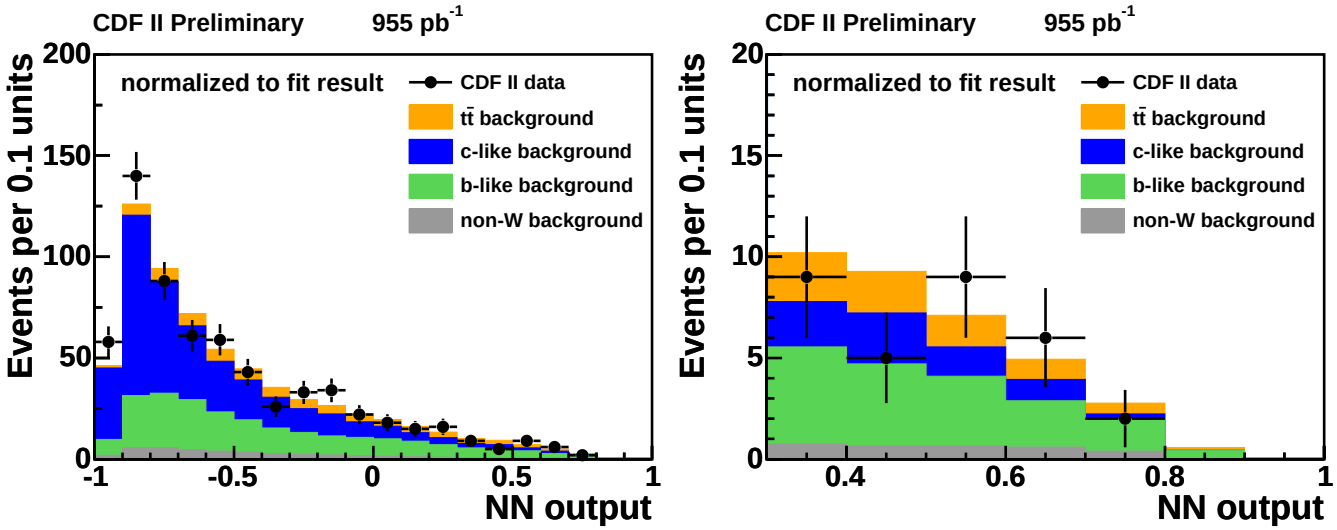

FIG. 14: Fit result versus data distribution. Left: in the entire NN output domain. Right: only the signal region. The normalization is the same in both histograms. Since the single-top contribution is zero, it is omitted in these histograms.

The resulting upper limit on the combined cross section is 2.6 pb at the 95% confidence level. The posterior probability density is shown in figure 15.

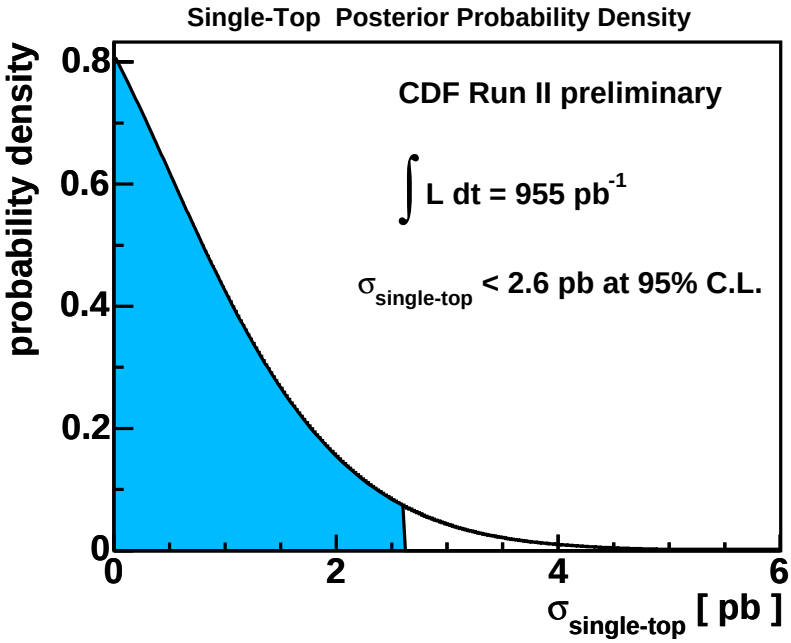

FIG. 15: Posterior probability density for the combined search using a neural network.

The observed Q-value is 9.13 which yields a p-value of 54.6%. The means that the observed data are well compatible with being a background fluctuation. In figure 16 we compare the observed Q-value to the expectation. The corresponding CL_{sb} value is 0.64%, that is the probability to observe this little single-top or less under the assumption of the predicted single-top cross section.

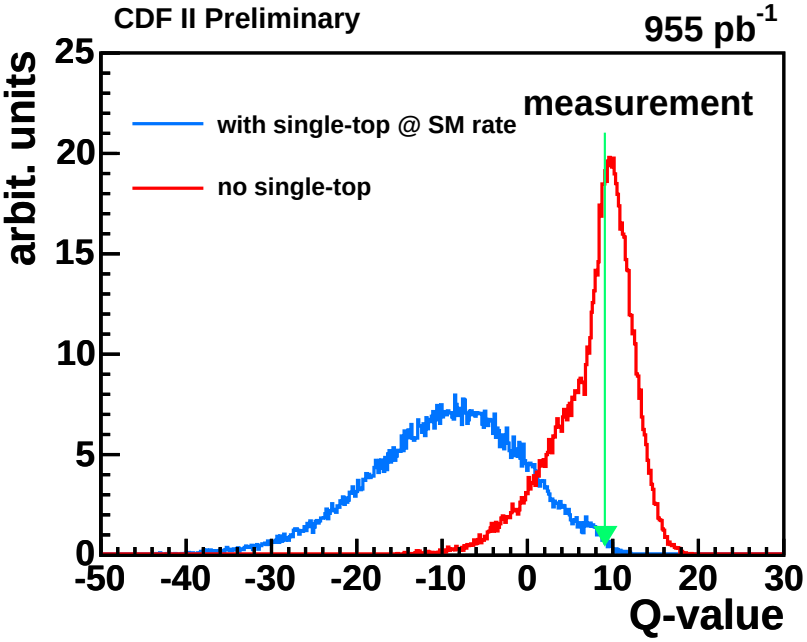

FIG. 16: Comparison of the observed Q-value to the expected distribution, (1) if single-top is present at the standard model rate, (2) if single-top is entirely absent.

X. BINNED LIKELIHOOD FIT TO DATA FOR SEPARATE SEARCH

As described above we apply a maximum likelihood fit to the network 2D output. The only difference in the likelihood function used is the generalization for two dimensions. The corresponding likelihood fit estimate of the cross sections at $\sigma_{t-ch} = 0.2^{+1.1}_{-0.2}$(stat. + syst.) pb and $\sigma_{s-ch} = 0.7^{+1.5}_{-0.7}$(stat. + syst.) pb is shown in figure 17. At the 95% confidence level the resulting upper limits on the t- and s-channel cross sections are 2.6 pb and 3.7 pb, respectively.

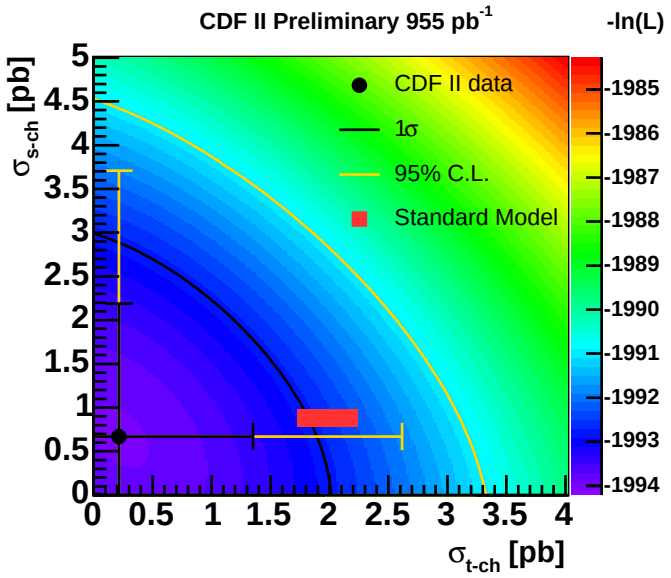

FIG. 17: The likelihood fit estimate for t- and s-channel cross section measurement. The contours of the 1σ uncertainty and the 95% C.L. are valid for both channels simultaneously. The error bars represent the 1σ uncertainty and the 95% C.L. of the given channel without assumptions on the other channel.

The observed Q-value of 2.94 yields a p-value of 21.9%. Figure 18 compares the observed Q-value to the expectation.

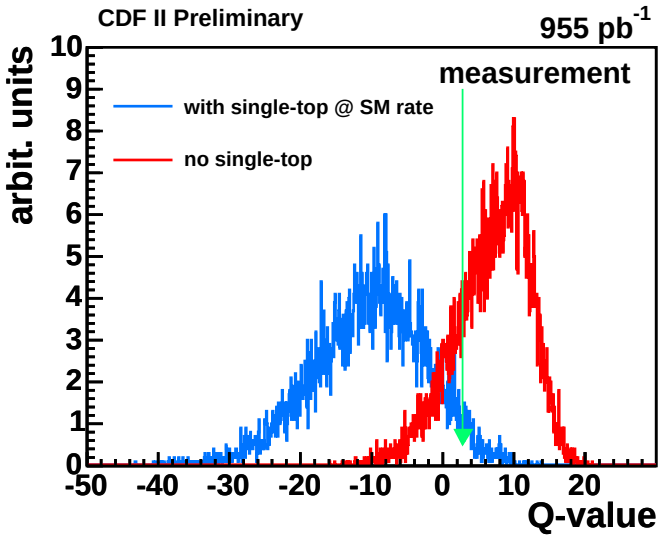

FIG. 18: Comparison of the observed Q-value for the 2D likelihood to the expected distribution, (1) if single-top is present at the standard model rate, (2) if single-top is entirely absent.

XI. CONCLUSIONS

We present a search of single-top quark production in a CDF II data set corresponding to 955 pb^{-1}, assuming a top quark mass of 175 GeV/c^2. We employ neural networks to construct a discriminant between single-top events and background events. In a combined search, where t- and s-channel single-top events are regarded as signal, we find an expected p-value of 0.5% which corresponds to a sensitivity of 2.6 σ. In data we observe no excess due to single-top events and compute a p-value of 52.5%, which indicates that the data are compatible with the background hypothesis only. We exclude a combined single-top cross section above 2.6 pb at the 95% confidence level.
For the separate search the observed t- and s-channel cross sections are

$$\sigma_{t-ch} = 0.2^{+1.1}_{-0.2}\,(\text{stat.} + \text{syst.})\ \text{pb}$$

$$\sigma_{s-ch} = 0.7^{+1.5}_{-0.7}\,(\text{stat.} + \text{syst.})\ \text{pb}$$

A summary of expected and observed upper limits at the 95% confidence level for the combined and the separate search is shown in figure 19.

CDF II Preliminary 955 pb^{-1}

Technique	Expected Limit	Observed Limit
NN combined	5.7 pb	2.6 pb
NN t-channel	3.8 pb	2.6 pb
NN s-channel	2.9 pb	3.7 pb

FIG. 19: Summary of expected and observed upper limits at the 95% confidence level.

[1] D. Acosta *et al.* (CDF Collaboration), *Phys. Rev.* **D65**, 091102 (2002); D. Acosta *et al.* (CDF Collaboration), *Phys. Rev.* **D69**, 052003 (2004); D. Acosta *et al.* (CDF Collaboration), *Phys. Rev.* **D71**, 012005 (2005).

[2] B. Abbott *et al.* (DØ Collaboration), *Phys. Rev.* **D63**, 031101 (2001); V. M. Abazov *et al.* (DØ Collaboration), *Phys. Lett.* **B517**, 282 (2001); V. M. Abazov *et al.* (DØ Collaboration), *Phys. Lett.* **B622**, 265 (2005); V. M. Abazov *et al.* (DØ Collaboration), hep-ex/0604020.

[3] B. W. Harris *et al.*, *Phys. Rev.* **D66**, 054024 (2002); Z. Sullivan, *Phys. Rev.* **D70**, 114012 (2004). J. Campbell, R.K. Ellis, F. Tramontano, *Phys. Rev.* **D70**, 094012 (2004).

[4] D. Acosta *et al.* (CDF Collaboration), *Phys. Rev.* **D71**, 052003.